Optimal Capacity of the Blume-Emery-Griffiths perceptron

D. Bollé* and I. Pérez Castillo†
Instituut voor Theoretische Fysica, K. U. Leuven, B-3001 Leuven, Belgium

G. M. Shim‡
Department of Physics, Chungnam National University, Yuseong, Daejeon 305-764 R. O. Korea

A Blume-Emery-Griffiths perceptron model is introduced and its optimal capacity is calculated within the replica-symmetric Gardner approach, as a function of the pattern activity and the imbedding stability parameter. The stability of the replica-symmetric approximation is studied via the analogue of the Almeida-Thouless line. A comparison is made with other three-state perceptrons.

PACS numbers: 02.50-r, 64.60.Cn, 75.10.Hk, 87.18.Sn

I. INTRODUCTION.

Recently an optimal Hamiltonian for a multistate network has been put forward [1], [2] on the basis of information theory by maximizing the mutual information content of the system. For a two-state network, this Hamiltonian equals the well-known Hopfield Hamiltonian extensively studied in the literature [3], [4]. For a three-state network one finds a Blume-Emery-Griffiths (BEG) spin-glass type Hamiltonian [5]. As spin-glasses these models have been studied for some time now. Thermodynamic as well as dynamic properties are discussed in the literature for disorder in both the quadratic and biquadratic interaction. Many references can be found in [6]. As a neural network model its study has been started only recently [2],[10]. But it turns out already that both the maximal capacity and the basin of attraction of this network are enlarged, at least for Hebb rule learning, in comparison with the standard three-state networks like, e.g., the Q-Ising spin-glass models.

A natural question is then whether these improved retrieval quality aspects are restricted to the use of the Hebb rule or whether they are intrinsic properties of the model. In the same context, a further question is then whether we can extract a perceptron type model with an optimal performance out of this BEG recurrent network. The perceptron is by now a well-known and standard model in theoretical studies and practical applications in connection with learning and generalization [3], [4], [7] - [9]. Consequently, a number of extensions including many-state, graded response and colored perceptrons have been formulated in the literature [11]-[18].

The aim of this work is precisely to introduce such a BEG-perceptron model and, in particular, to study its Gardner optimal capacity. Although the method for doing that is standard and well-know by now [19],[20] its generalization to the problem at hand is highly non-trivial. Nevertheless we have succeeded in obtaining a closed expression for the replica symmetric approximation to the Gardner optimal capacity.

The paper is organized as follows. In section 2 we recall the BEG Hamiltonian and define the BEG perceptron model. Section 3 presents a closed analytic formula for the replica-symmetric Gardner capacity of this model and studies its behaviour as a function of the imbedding constant and the activity. Comparisons with other three-state perceptrons are made. In section 4 the stability of the replica symetric solution is studied using an extension of the de Almeida-Thouless analysis. The analytic form of the two replicon eigenvalues is obtained. Stability is found to be broken for smaller values of the activity and for very small imbedding stabilities. Section 5 presents some concluding remarks. In the appendices further technical explanations are given.

II. THE BEG PERCEPTRON.

Consider a neural network consisting of N neurons which can take values $\sigma_i, i = 1, \ldots, N$ from the discrete set $\mathcal{S} \equiv \{-1, 0, +1\}$. The p patterns to be stored in this network are supposed to be a collection of independent and identically distributed random variables (i.i.d.r.v.), ξ_i^μ, $\mu = 1, \ldots, p$ with a probability distribution

$$p(\xi_i^\mu) = \frac{a}{2}\delta(\xi_i^\mu - 1) + \frac{a}{2}\delta(\xi_i^\mu + 1) + (1-a)\delta(\xi_i^\mu) \quad (1)$$

with a the activity of the patterns so that

$$\lim_{N\to\infty} \frac{1}{N}\sum_i (\xi_i^\mu)^2 = a. \quad (2)$$

Given the network configuration at time t, $\boldsymbol{\sigma}_N \equiv \{\sigma_j(t)\}, j = 1, \ldots, N$, the following dynamics is considered. The configuration $\boldsymbol{\sigma}_N(0)$ is chosen as input. The neurons are updated according to the stochastic parallel spin-flip dynamics defined by the transition probabilities

$$\Pr\left(\sigma_i(t+1) = s' \in \mathcal{S} | \boldsymbol{\sigma}_N(t)\right) = \frac{\exp[-\beta\epsilon_i(s'|\boldsymbol{\sigma}_N(t))]}{\sum_{s\in\mathcal{S}}\exp[-\beta\epsilon_i(s|\boldsymbol{\sigma}_N(t))]}\,. \quad (3)$$

*Electronic address: Desire.Bolle@fys.kuleuven.ac.be
†Electronic address: Isaac.Perez@fys.kuleuven.ac.be
‡Electronic address: gmshim@cnu.ac.kr

Here the energy potential $\epsilon_i[s|\boldsymbol{\sigma}_N(t)]$ is defined by

$$\epsilon_i[s|\boldsymbol{\sigma}_N(t)] = -sh_i(\boldsymbol{\sigma}_N(t)) - s^2\theta_i(\boldsymbol{\sigma}_N(t)), \qquad (4)$$

where the following local fields in neuron i carry all the information

$$h_{N,i}(t) = \sum_{j\neq i} J_{ij}\sigma_j(t), \quad \theta_{N,i}(t) = \sum_{j\neq i} K_{ij}\sigma_j^2(t) \qquad (5)$$

with the obvious shorthand notation for the local fields. For synaptic couplings J_{ij} and K_{ij} of the Hebb-type

$$J_{ij} = \frac{1}{a^2N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu \qquad (6)$$

$$K_{ij} = \frac{1}{a^2(1-a)^2N}\sum_{\mu=1}^{p}((\xi_i^\mu)^2-a)((\xi_j^\mu)^2-a) \qquad (7)$$

the corresponding neural network Hamiltonian

$$H = -\frac{1}{2}\sum_{i\neq j} J_{ij}\sigma_i\sigma_j - \frac{1}{2}\sum_{i\neq j}K_{ij}\sigma_i^2\sigma_j^2\,, \qquad (8)$$

has been discussed recently [2]. It has been found that the capacity and basin of attraction has been enlarged in comparison with other three-state networks.

We would like to understand whether these better retrieval quality is an intrinsic property of the model. Therefore, we want to answer the following question: given the set of p patterns specified above, is there a network (the best possible network of the BEG-type) which has these patterns as fixed points of the deterministic form of the dynamics considered above? At zero temperature the updating rule of this dynamics (3)-(4) is equivalent to the gain function

$$\begin{aligned}\sigma_i(t+1) &= \mathrm{sign}(h_{N,i}(t))\Theta(|h_{N,i}(t)|+\theta_{N,i}(t)) \\ &\equiv \mathrm{g}(h_{N,i}(t),\theta_{N,i}(t))\end{aligned} \qquad (9)$$

with Θ the Heaviside function. Considering the perceptron architecture (N inputs with couplings J_j and K_j and 1 output) we say that a given pattern, $\xi_i^\mu, i=1,\ldots,N$, is stored if there exists a corresponding output ξ_0^μ

$$\xi_0^\mu = g(h^\mu,\theta^\mu) \qquad (10)$$

with

$$h^\mu = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}J_j\xi_j^\mu \qquad \theta^\mu = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}K_j(\xi_j^\mu)^2\,, \qquad (11)$$

and $\{\mathbf{J},\mathbf{K}\}\equiv\{J_j,K_j\}$ denoting the configurations in the space of interactions. The factor $N^{-1/2}$ is introduced to have the weights J_j and K_j of order unity.

The aim is then to determine the maximal number of patterns, p, that can be stored in the perceptron, in other words to find the maximal value of the loading $\alpha = p/N$ for which couplings satisfying (10)-(11) can still be found. Following a Gardner-type analysis [19] the fundamental quantity that we want to calculate is then the volume fraction of weight space given by

$$V = \int d\mathbf{J}d\mathbf{K}\rho(\mathbf{J},\mathbf{K})\prod_{\mu=1}^{p}\chi_{\xi_0^\mu}(h^\mu,\theta^\mu;\kappa) \qquad (12)$$

with the characteristic function

$$\begin{aligned}\chi_{\xi_0^\mu}(h^\mu,\theta^\mu;\kappa) &= \delta_{\xi_0^\mu,g(h^\mu,\theta^\mu)} \\ &= (\xi_0^\mu)^2\Theta(|h^\mu|+\theta^\mu-\kappa)\Theta(\xi_0^\mu h^\mu-\kappa) \\ &\quad +(1-(\xi_0^\mu)^2)\Theta(-|h^\mu|-\theta^\mu-\kappa)\end{aligned} \qquad (13)$$

where κ is the imbedding stability parameter measuring the size of the basin of attraction for the μ-th pattern and $\rho(\mathbf{J},\mathbf{K})$ is the following normalization factor assuming spherical constraints for the couplings

$$\rho(\mathbf{J},\mathbf{K}) = \frac{\delta(\mathbf{J}\cdot\mathbf{J}-N)\delta(\mathbf{K}\cdot\mathbf{K}-N)}{\int_{-\infty}^{\infty}d\mathbf{J}d\mathbf{K}\delta(\mathbf{J}\cdot\mathbf{J}-N)\delta(\mathbf{K}\cdot\mathbf{K}-N)}\,. \qquad (14)$$

In order to perform the average over the disorder in the input patterns and the corresponding output we employ the replica technique to evaluate the entropy per site

$$v = \lim_{N\to\infty}\frac{1}{N}\langle\!\langle \ln V\rangle\!\rangle \qquad (15)$$

where $\langle\!\langle\cdots\rangle\!\rangle$ denotes an average over the statistics of inputs $\{\xi_j^\mu\}$ and outputs $\{\xi_0^\mu\}$, recalling (1).

III. REPLICA SYMMETRIC ANALYSIS

In the replica approach the entropy per site v is computed via the expression

$$v = \lim_{N\to\infty}\lim_{n\to 0}\frac{1}{nN}\big(\langle\!\langle V^n\rangle\!\rangle - 1\big) = \lim_{N\to\infty}\lim_{n\to 0}\frac{1}{nN}\ln\langle\!\langle V^n\rangle\!\rangle \qquad (16)$$

where V^n is the n-times replicated fractional volume

$$\langle\!\langle V^n\rangle\!\rangle \propto \int\Big[\prod_{\alpha=1}^{n}d\mathbf{J}^\alpha d\mathbf{K}^\alpha\delta\Big(\mathbf{J}^\alpha\cdot\mathbf{J}^\alpha-N\Big)\delta\Big(\mathbf{K}^\alpha\cdot\mathbf{K}^\alpha-N\Big)\Big]\Big\langle\!\!\Big\langle\prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\chi_{\xi_0^\mu}(h_\mu^\alpha,\theta_\mu^\alpha;\kappa)\Big\rangle\!\!\Big\rangle \qquad (17)$$

whereby we can forget, since the couplings are continuous, about constant terms such as the denominator in (14). The calculation then proceeds in a standard way although the technical details are much more complicated. For a short account we refer to Appendix A. Here we restrict ourselves to the following important remarks. The main order parameters appearing in the calculation are

$$q_{\alpha\beta} = \frac{1}{N}\mathbf{J}^\alpha \cdot \mathbf{J}^\beta, \quad r_{\alpha\beta} = \frac{1}{N}\mathbf{K}^\alpha \cdot \mathbf{K}^\beta, \quad \alpha < \beta$$
$$L^\alpha = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} K_j^\alpha, \quad \forall \alpha. \tag{18}$$

Of course, in the replica symmetric (RS) approximation we are focussing upon here, $q_{\alpha\beta} = q, r_{\alpha\beta} = r, L^\alpha = L$. The first two order parameters are the overlaps between two distinct replicas for the couplings $\mathbf{J}$ and $\mathbf{K}$, the third one arises from the fact that the dynamics (9) and, hence, also the characteristic function (13), contains a second field θ, quadratic in the patterns. We remark that it describes the relative importance of the active versus the non-active neurons. Actually, in the calculation aL will be the important quantity with a the second moment of the pattern distribution, i.e., the pattern activity.

The RS optimal Gardner capacity is obtained when the overlap order parameters q and r go to 1. It is clear that these limits have to be taken simultaneously but, in general, their rate of convergence could be different. Therefore, we introduce $(1-r) = \gamma(1-q)$ where γ is a new parameter which one also needs to extremize. We expect this parameter γ to depend on the pattern distribution through the activity a.

Pursuing this approach then leads to

$$\alpha_{RS}(a,\kappa) = -\underset{L,\gamma}{\text{extr}} \lim_{q\to 1} \frac{1+1/\gamma}{2(1-q) g_1^{RS}(q,\gamma,L)} \tag{19}$$

where $g_1^{RS}(q,\gamma,L)$ reads

$$g_1^{RS}(q,\gamma,L) = \int \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - l) \left\langle\!\!\left\langle \ln \int_{\Omega_\xi} \frac{dh}{\sqrt{2\pi(1-q)}} \frac{d\theta}{\sqrt{2\pi(1-q)}} \exp\left[-\frac{(h-h_0)^2 + (\theta-\theta_0)^2}{2(1-q)}\right] \right\rangle\!\!\right\rangle_{\xi_0} \tag{20}$$

with $l \equiv aL/\sqrt{a(1-a)}$, where $\mathcal{D}(ax+b) = (2\pi)^{-1/2} a dx \exp[(-1/2)(ax+b)^2]$ and where the integration region Ω_ξ is determined by the Heaviside functions appearing in the characteristic function $\chi_\xi(\sqrt{a}h, \sqrt{\gamma a(1-a)}\theta; \kappa)$ defined in (13). The expression (20) for the function g_1^{RS} suggests that an asymptotic expansion to compute the limit $q \to 1$ is possible. Indeed, after some tedious algebra (see Appendix B) we find for this limit

$$\begin{aligned} g_1^{RS}(q,\gamma,L) &= -\frac{a}{2(1-q)} \sum_{i=1}^{3} \int_{\mathcal{R}_i} \mathcal{D}(h_0 + \kappa/\sqrt{a}\mathcal{D}(\sqrt{\gamma}\theta_0 - l)\, d_{min}^{\mathcal{R}_i}(h_0,\theta_0) \\ &\quad -\frac{(1-a)}{2(1-q)} \sum_{i=1}^{3} \int_{\mathcal{R}'_i} \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - u)\, d^{\mathcal{R}'_i}(h_0,\theta_0) + o(1/(1-q)) \end{aligned} \tag{21}$$

with $u \equiv (aL+\kappa)/\sqrt{a(1-a)}$. The integration regions read

$$\mathcal{R}_1 = \begin{cases} h_0 < 0 \\ \theta_0 > 0 \end{cases} \tag{22}$$
$$\mathcal{R}_2 = \begin{cases} h_0\gamma' < \theta_0 < 0 \\ h_0 < 0 \end{cases} \tag{23}$$
$$\mathcal{R}_3 = \begin{cases} \theta_0 < 0 \\ \theta_0/\gamma' < h_0 < -\theta_0\gamma' \end{cases} \tag{24}$$
$$\mathcal{R}'_1 = \begin{cases} h_0 > 0 \\ -h_0/\gamma' < \theta_0 < \gamma' h_0 \end{cases} \tag{25}$$
$$\mathcal{R}'_2 = \begin{cases} -\theta_0/\gamma' < h_0 < \theta_0/\gamma' \\ \theta_0 > 0 \end{cases} \tag{26}$$
$$\mathcal{R}'_3 = \begin{cases} h_0 < 0 \\ h_0/\gamma' < \theta_0 < -\gamma' h_0 \end{cases} \tag{27}$$

and the corresponding integrands are given by

$$d_{min}^{\mathcal{R}_1} = h_0^2 \tag{28}$$
$$d_{min}^{\mathcal{R}_2} = h_0^2 + \theta_0^2 \tag{29}$$
$$d_{min}^{\mathcal{R}_3} = \frac{1}{1+(\gamma')^2}\left(h_0 + \gamma'\theta_0\right)^2 \tag{30}$$
$$d_{min}^{\mathcal{R}'_1} = \frac{1}{1+(\gamma')^2}\left(h_0 + \gamma'\theta_0\right)^2 \tag{31}$$
$$d_{min}^{\mathcal{R}'_2} = h_0^2 + \theta_0^2 \tag{32}$$
$$d_{min}^{\mathcal{R}'_3} = \frac{1}{1+(\gamma')^2}\left(h_0 - \gamma'\theta_0\right)^2 \tag{33}$$

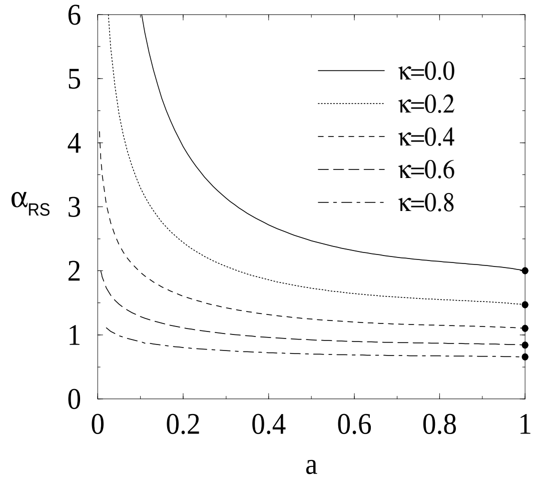

FIG. 1: The optimal capacity α_{RS} as a function of the pattern activity a for several values of the stability constant κ. The dots at $a = 1$ refer to the optimal capacity of the two-state perceptron.

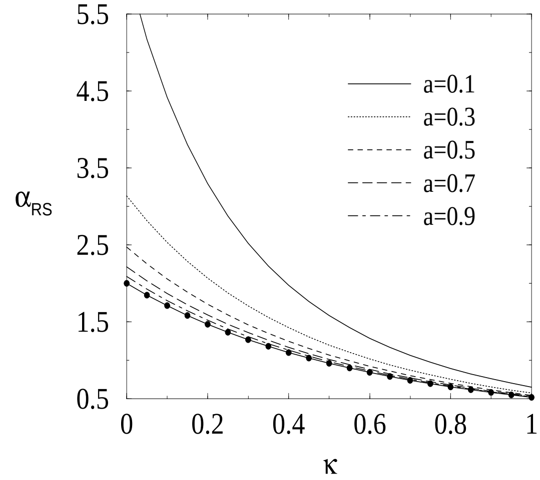

FIG. 2: The optimal capacity α_{RS} as a function of the stability κ for several values of the pattern activity a. The straight-dotted line corresponds to the optimal capacity of the two-state perpectron.

with $\gamma' \equiv \sqrt{\gamma(1-a)}$ and where we remark that the d_{min} are minimal distances between a point in the different integration regions $\mathcal{R}_i, \mathcal{R}'_i, i = 1, 2, 3$ and the border of Ω_ξ(see Appendix B). This may allow for a possible geometrical interpretation of the Gardner optimal capacity in the space of local fields as it has been suggested for the Q-state clock model in [21].

After inserting (21)-(33) in (19) and extremizing numerically with respect to L and γ, we find the results presented in figures 1-2. In fig. 1 the capacity α_{RS} versus the activity a is shown for several values of the imbedding stability constant κ. For bigger κ, the capacity becomes, of course, smaller. For $a = 1$, i.e., binary patterns, we find back the original Gardner results, as we do in fig. 2 showing α_{RS} as a function of κ for several values of a. Smaller activity indicating a growing presence of zero-state neurons leads to bigger capacities. Of course, this does not mean a priori that also the information content of the system is increased. For completeness, we remark that the parameters $l = aL/\sqrt{a(1-a)}$ and γ that we have extremized over, depend rather strongly but smoothly on the pattern activity. For $a = 1$ we find back the two-state perceptron value for L, i.e. $L = 0 (l = \infty)$, and $\gamma = \infty$. Finally, in order to have an idea about the information stored into the network we plot in fig. 3 the information content per neuron

$$I(a) = -\frac{\alpha_{RS}}{\ln 3}\left[a \ln(\frac{a}{2}) + (1-a)\ln(1-a)\right]. \qquad (34)$$

For $a = 1$ our result is again consistent with the simple perceptron result [19]. Comparing with other three-state neuron perceptron models we recall that for $\kappa = 0$ and uniform patterns the $Q = 3$ Ising perceptron can maximally reach an optimal capacity equal to 1.5, depending on the separation between the plateaus of the gain function (see [14], [15]) for the precise details) and the $Q = 3$ clock and Potts model both reach an optimal capacity of 2.40 [12],[21] while the value for the BEG perceptron found here is 2.24. Here we have to recall that the $Q = 3$ Ising perceptron and the BEG perceptron have the same topology structure in the neurons, whereas the $Q = 3$ clock and Potts models have a different topology.

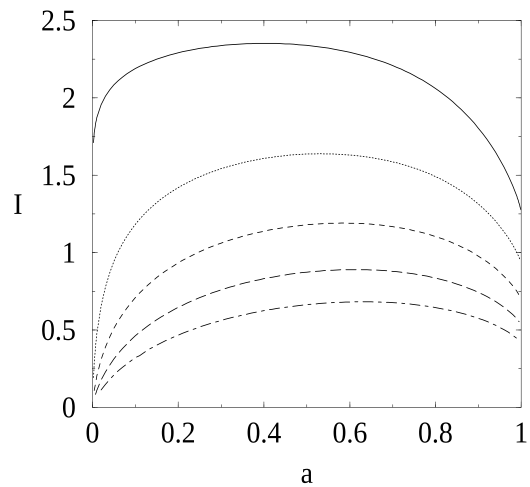

FIG. 3: The information content per neuron, I, as a function of a for $\kappa = 0; 0.2; 0.4; 0.6; 0.8$ (from top to bottom)

IV. STABILITY OF THE REPLICA SYMMETRIC SOLUTION

From the work of Gardner [19] we know that for the binary neuron perceptron the RS solution is marginally stable against RS breaking (RSB) fluctuations. From the work on multi-state Q-Ising neurons [24] we know that

the RS solution may be stable or unstable depending on the gain parameter, the number of spin states and the distribution of the patterns. Furthermore, in general, increasing the imbedding stability parameter κ lowers the capacity and enhances the stability against RSB. Using these results for the $Q = 3$ spin states as a guide we also expect breaking for the BEG perceptron model at hand. To confirm this and find out the precise interval of a values where breaking occurs, we generalize the de Almeida-Thouless analysis [22], [23].

First, the hessian matrix associated with the function Φ, eq. (A8), is computed, and then the eigenvalues are determined. As usual, two types of eigenvalues are found: longitudinal eigenvalues describing fluctuations within RS and transverse eigenvalues describing stability against RSB. We find *four* transversal eigenvalues each with degeneracy $\frac{1}{2}n(n-3)$. In the limit $q \to 1$ they can be calculated explicitly in terms of the minimal distances occuring in (28)-(33). The result reads (for more details we refer to Appendix C)

$$\lambda_+ = \frac{1}{2}(\Delta_q + \Delta_r) + \frac{1}{2}\sqrt{(\Delta_q - \Delta_r)^2 + 4\Delta_c^2} \tag{35}$$

$$\lambda_- = \frac{1}{2}(\Delta_q + \Delta_r) - \frac{1}{2}\sqrt{(\Delta_q - \Delta_r)^2 + 4\Delta_c^2} \tag{36}$$

$$\tau_+ = \frac{1}{2(\Delta_c^2 - \Delta_q\Delta_r)}\Big\{\Delta_q + \Delta_r + (\Delta_{\hat{q}} + \Delta_{\hat{r}})(\Delta_c^2 - \Delta_q\Delta_r) + \sqrt{4\Delta_c^2 + \big[\Delta_q - \Delta_r + (\Delta_{\hat{q}} - \Delta_{\hat{r}})(\Delta_q\Delta_r - \Delta_c^2)\big]^2}\Big\} \tag{37}$$

$$\tau_- = \frac{1}{2(\Delta_c^2 - \Delta_q\Delta_r)}\Big\{\Delta_q + \Delta_r + (\Delta_{\hat{q}} + \Delta_{\hat{r}})(\Delta_c^2 - \Delta_q\Delta_r) - \sqrt{4\Delta_c^2 + \big[\Delta_q - \Delta_r + (\Delta_{\hat{q}} - \Delta_{\hat{r}})(\Delta_q\Delta_r - \Delta_c^2)\big]^2}\Big\} \tag{38}$$

with the Δ's given by

$$\begin{aligned}
\Delta_q &= \frac{a\alpha_{RS}}{(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}_i}\mathcal{D}(h_0+\kappa/\sqrt{a}, \sqrt{\gamma}\theta_0 - t)\Big\{\frac{1}{2}\frac{\partial^2}{\partial h_0^2}d_{min}^{\mathcal{R}_i}(h_0,\theta_0)\Big\}^2 \\
&+\frac{(1-a)\alpha_{RS}}{(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}'_i}\mathcal{D}(h_0, \sqrt{\gamma}\theta_0 - u)\Big\{\frac{1}{2}\frac{\partial^2}{\partial h_0^2}d_{min}^{\mathcal{R}'_i}(h_0,\theta_0)\Big\}^2 + o(1/(1-q)) \qquad (39)\\
\Delta_r &= \frac{a\alpha_{RS}}{\gamma 2(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}_i}\mathcal{D}(h_0+\kappa/\sqrt{a}, \sqrt{\gamma}\theta_0 - t)\Big\{\frac{1}{2}\frac{\partial^2}{\partial \theta_0^2}d_{min}^{\mathcal{R}_i}(h_0,\theta_0)\Big\}^2 \\
&+\frac{(1-a)\alpha_{RS}}{\gamma 2(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}'_i}\mathcal{D}(h_0, \sqrt{\gamma}\theta_0 - u)\Big\{\frac{1}{2}\frac{\partial^2}{\partial \theta_0^2}d_{min}^{\mathcal{R}'_i}(h_0,\theta_0)\Big\}^2 + o(1/(1-q)) \qquad (40)\\
\Delta_c &= \frac{a\alpha_{RS}}{\gamma(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}_i}\mathcal{D}(h_0+\kappa/\sqrt{a}, \sqrt{\gamma}\theta_0 - t)\Big\{\frac{1}{2}\frac{\partial^2}{\partial h_0\partial\theta_0}d_{min}^{\mathcal{R}_i}(h_0,\theta_0)\Big\}^2 \\
&+\frac{(1-a)\alpha_{RS}}{\gamma(1-q)^2}\sum_{i=1}^{3}\int_{\mathcal{R}'_i}\mathcal{D}(h_0, \sqrt{\gamma}\theta_0 - u)\Big\{\frac{1}{2}\frac{\partial^2}{\partial h_0\partial\theta_0}d_{min}^{\mathcal{R}'_i}(h_0,\theta_0)\Big\}^2 + o(1/(1-q)) \qquad (41)\\
\Delta_{\hat{q}} &= (1-q)^2 \qquad (42)\\
\Delta_{\hat{r}} &= (1-r)^2 = \gamma^2(1-q)^2\,. \qquad (43)
\end{aligned}$$

Then *two* replicon eigenvalues, λ_{R_1} and λ_{R_2}, can be defined as

$$\lambda_{R_1} = \lambda_+\tau_- \qquad \lambda_{R_2} = \lambda_-\tau_+ \tag{44}$$

Stability of the RS solution requires that both $\lambda_{R_1}, \lambda_{R_2} < 0$. In fig. 4-6 we present the numerical results concerning the stability analysis. In fig. 4 the first replicon eigenvalue λ_{R_1} is shown as a function of a for several values of κ. It is seen that for small values of κ this eigenvalue becomes positive for smaller values of a and

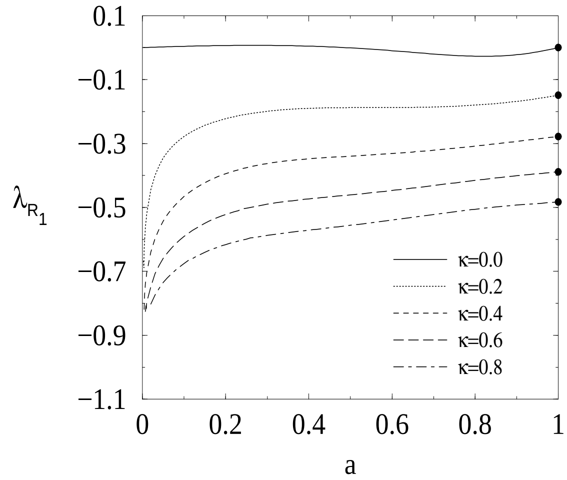

FIG. 4: The first replicon eigenvalue λ_{R_1} as a function of a for several values of κ. The dots at $a = 1$ refer to the optimal capacity of the two-state perceptron.

hence replica symmetry is broken. We remark that for $a = 1$ our results are consistent with those of Gardner [19]. Fig. 5 presents a closer view of this for $\kappa = 0$. For $0 < a \leq 0.48(8)$ the RS solution is unstable. Storing only zero-state spins, $a = 0$, or binary spins $a = 1$ leads to marginal stability. As a first explanation one could remark that for increasing a, allowing more $\pm$ states, the disorder is increased up to about a uniform distribution of patterns, $a = 2/3$. It is clear that for bigger κ, the stability against RSB increases. In fact for $\kappa > 0.0061$ already no more breaking occurs. Finally, fig. 6 shows

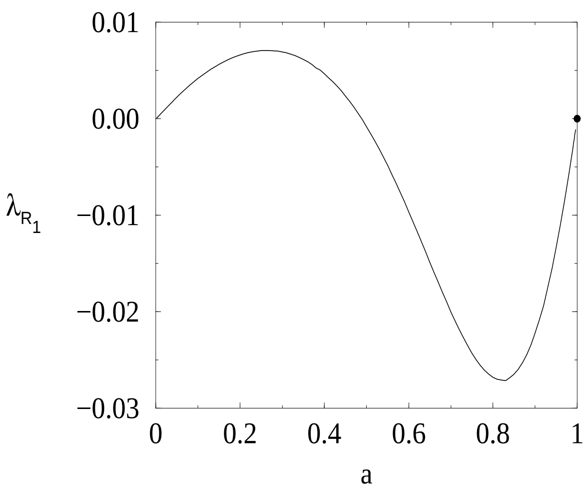

FIG. 5: The first replicon eigenvalue λ_{R_1} as a function of a for $\kappa = 0$ on a different scale. RSB occurs for smaller values of a.

that λ_{R_2} is always negative and, hence, plays no role in the breaking of the RS stability.

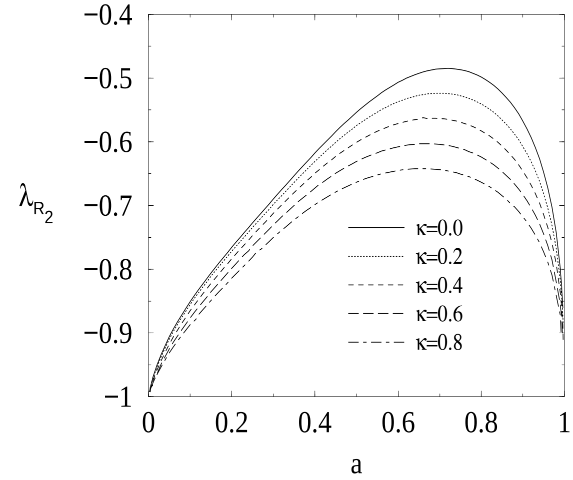

FIG. 6: The second replicon eigenvalue λ_{R_2} as a function of a for several values of κ.

V. CONCLUDING REMARKS.

In this work we have introduced a perceptron model based upon the recently studied Blume-Emery-Griffiths neural network, containing ternary neurons. We have obtained an analytic formula for the replica symmetric optimal Gardner capacity. For the imbedding stability constant equal to zero and uniform patterns, e.g., we find a bigger optimal capacity, $\alpha_{RS} = 2.24$, than the one for the $Q = 3$ Ising perceptron, $\alpha_{RS} = 1.5$, which has the same topology structure for the neurons. Since, in general, perceptrons turn out to be very useful models in connection with learning and generalization this is an interesting observation. It is also consistent with earlier results derived for the Hebb rule.

We have studied the stability of the replica-symmetric solution by generalizing the de Almeida-Thouless analysis and deriving an analytic expression for the *two* replicon eigenvalues that play a role in the Gardner limit. Breaking only occurs for small activities and very small imbedding constants, $\kappa < 0.0061$. This is consistent with the stability results found for the $Q = 3$ Ising perceptrons.

These results strenghten the idea that the better retrieval properties found for the Blume-Emery-Griffiths model in comparison with the $Q = 3$ Ising model are not restricted to the specific Hebb rule but are intrinsic to the model.

Acknowledgments

We would like to thank Dr. P. Kozlowski and T. Verbeiren for critical discussions. One of us (IPC) especially thanks J. van Mourik and J. Gheerardyn for useful remarks. This work has been supported in part by the Fund of Scientific Research, Flanders-Belgium.

APPENDIX A: REPLICA ANALYSIS AND REPLICA SYMMETRIC ANSATZ.

In this appendix we outline the main steps in the calculation of the n-times replicated volume (17) extending [19] to the case at hand. In order to perform the quenched average we use the δ-function representation

$$1=\int_{-\infty}^{\infty}\frac{dh_\mu^\alpha d\widehat{h}_\mu^\alpha}{2\pi}\exp\left[i\widehat{h}_\mu^\alpha\Big(h_\mu^\alpha-\frac{1}{\sqrt{N}}\sum_{j=1}^{N}J_j^\alpha\boldsymbol{\xi}_j^\mu\Big)\right] \tag{A1}$$

$$1=\int_{-\infty}^{\infty}\frac{d\theta_\mu^\alpha d\widehat{\theta}_\mu^\alpha}{2\pi}\exp\left[i\widehat{\theta}_\mu^\alpha\Big(\theta_\mu^\alpha-\frac{1}{\sqrt{N}}\sum_{j=1}^{N}K_j^\alpha(\boldsymbol{\xi}_j^\mu)^2\Big)\right] \tag{A2}$$

to take the local fields out of the characteristic function and obtain

$$\left\langle\!\!\left\langle\prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\chi_{\xi_0^\mu}(h_\mu^\alpha,\theta_\mu^\alpha;\kappa)\right\rangle\!\!\right\rangle=\int\Big[\prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\frac{d\theta_\mu^\alpha d\widehat{\theta}_\mu^\alpha}{2\pi}\frac{dh_\mu^\alpha d\widehat{h}_\mu^\alpha}{2\pi}\Big]\exp\Big[i\sum_{\alpha=1}^{n}\sum_{\mu=1}^{p}\big(\widehat{h}_\mu^\alpha h_\mu^\alpha+\widehat{\theta}_\mu^\alpha\theta_\mu^\alpha\big)\Big]$$
$$\left\langle\!\!\left\langle\prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\exp\Big[-\frac{i\widehat{h}_\mu^\alpha}{\sqrt{N}}\sum_{j=1}^{N}J_j^\alpha\xi_j^\mu-\frac{i\widehat{\theta}_\mu^\alpha}{\sqrt{N}}\sum_{j=1}^{N}K_j^\alpha(\xi_j^\mu)^2\Big]\right\rangle\!\!\right\rangle_{\xi_i^\mu}\left\langle\!\!\left\langle\prod_{\alpha=1}^{n}\prod_{\mu=1}^{p}\chi_{\xi_0^\mu}(h_\mu^\alpha,\theta_\mu^\alpha;\kappa)\right\rangle\!\!\right\rangle_{\xi_o^\mu}. \tag{A3}$$

Introducing the order parameters (18)and their conjugate variables, and using the identities

$$1=\int_{-\infty}^{\infty}\prod_{\alpha<\beta}\frac{dq_{\alpha\beta}d\widehat{q}_{\alpha\beta}}{2\pi i/N}\exp\left[\widehat{q}_{\alpha\beta}\Big(Nq_{\alpha\beta}-\mathbf{J}^\alpha\cdot\mathbf{J}^\beta\Big)\right] \tag{A4}$$

$$1=\int_{-\infty}^{\infty}\prod_{\alpha<\beta}\frac{dr_{\alpha\beta}d\widehat{r}_{\alpha\beta}}{2\pi i/N}\exp\left[\widehat{r}_{\alpha\beta}\Big(Nr_{\alpha\beta}-\mathbf{K}^\alpha\cdot\mathbf{K}^\beta\Big)\right] \tag{A5}$$

$$1=\int_{-\infty}^{\infty}\prod_{\alpha=1}^{n}\frac{dL^\alpha d\widehat{L}^\alpha}{2\pi/\sqrt{N}}\exp\left[i\widehat{L}^\alpha\Big(\sqrt{N}L^\alpha-\sum_{j=1}^{N}K_j^\alpha\Big)\right] \tag{A6}$$

allows us to express the replicated fractional volume as an integral over them, viz.

$$\langle\!\langle V^n\rangle\!\rangle\propto\int_{-\infty}^{\infty}\Big[\prod_{\alpha=1}^{n}\frac{dL^\alpha d\widehat{L}^\alpha}{2\pi/\sqrt{N}}\Big]\Big[\prod_{\alpha=1}^{n}\frac{d\widehat{E}^\alpha}{4\pi i}\frac{d\widehat{F}^\alpha}{4\pi i}\Big]\Big[\prod_{\alpha<\beta}\frac{dq_{\alpha\beta}d\widehat{q}_{\alpha\beta}}{2\pi i/N}\frac{dr_{\alpha\beta}d\widehat{r}_{\alpha\beta}}{2\pi i/N}\Big]\exp\Big[N\Phi\Big] \tag{A7}$$

with Φ given by

$$\Phi=\alpha G_1(q_{\alpha\beta},r_{\alpha\beta},L^\alpha)+G_2(\widehat{Q}_{\alpha\beta},\widehat{R}_{\alpha\beta},\widehat{L}^\alpha)+G_3(q_{\alpha\beta},r_{\alpha\beta},\widehat{Q}_{\alpha\beta},\widehat{R}_{\alpha\beta}) \tag{A8}$$

where

$$\begin{aligned}G_1=\;&\ln\int_{-\infty}^{\infty}\Big[\prod_{\alpha=1}^{n}\frac{d\theta^\alpha d\widehat{\theta}^\alpha}{2\pi}\frac{dh^\alpha d\widehat{h}^\alpha}{2\pi}\Big]\exp\Big[i\sum_{\alpha=1}^{n}(\widehat{h}^\alpha h^\alpha+\widehat{\theta}^\alpha\theta^\alpha)-ia\sum_{\alpha=1}^{n}\widehat{\theta}^\alpha L^\alpha-\frac{a}{2}\sum_{\alpha,\beta=1}^{n}\widehat{h}^\alpha\widehat{h}^\beta q_{\alpha\beta}\\&-\frac{a(1-a)}{2}\sum_{\alpha,\beta=1}^{n}\widehat{\theta}^\alpha\widehat{\theta}^\beta r_{\alpha\beta}\Big]\left\langle\!\!\left\langle\prod_{\alpha=1}^{n}\chi_\xi(h^\alpha,\theta^\alpha;\kappa)\right\rangle\!\!\right\rangle_{\xi_o}\end{aligned} \tag{A9}$$

$$G_2=\ln\int_{-\infty}^{\infty}\Big[\prod_{\alpha=1}^{n}dJ^\alpha dK^\alpha\Big]\exp\Big[-\frac{1}{2}\sum_{\alpha,\beta=1}^{n}\Big(\widehat{Q}_{\alpha\beta}J^\alpha J^\beta+\widehat{R}_{\alpha\beta}K^\alpha K^\beta\Big)-i\sum_{\alpha=1}^{n}\widehat{L}^\alpha K^\alpha\Big] \tag{A10}$$

$$G_3=\frac{1}{2}\sum_{\alpha,\beta=1}^{n}\Big(\widehat{Q}_{\alpha\beta}Q_{\alpha\beta}+\widehat{R}_{\alpha\beta}R_{\alpha\beta}\Big) \tag{A11}$$

and

$$\widehat{Q}_{\alpha\beta}=\widehat{E}^\alpha\delta_{\alpha\beta}+\widehat{q}_{\alpha\beta}(1-\delta_{\alpha\beta}) \tag{A12}$$
$$\widehat{R}_{\alpha\beta}=\widehat{F}^\alpha\delta_{\alpha\beta}+\widehat{r}_{\alpha\beta}(1-\delta_{\alpha\beta}) \tag{A13}$$
$$Q_{\alpha\beta}=\delta_{\alpha\beta}+q_{\alpha\beta}(1-\delta_{\alpha\beta}) \tag{A14}$$
$$R_{\alpha\beta}=\delta_{\alpha\beta}+r_{\alpha\beta}(1-\delta_{\alpha\beta})\,. \tag{A15}$$

We remark that the δ-function representation of the local fields has allowed us to perform the calculations until this

point without using an explicit form for the characteristic function $\chi_\xi(h^\alpha, \theta^\alpha; \kappa)$. Using the RS ansatz Φ can be simplified further and the saddle-point equations for $\hat{Q}, Q, \hat{R}, R$ become algebraic so that they can be solved explicitly, leading to the result (19)-(20).

APPENDIX B: $q \to 1$ LIMIT

In order to compute the asymptotic expansion of (20) we proceed as follows. We split the integral over (h_0, θ_0) into two parts, i.e., Ω_ξ determined by the Heaviside function in χ_ξ, and its complement $\mathcal{C}(\Omega_\xi)$. The first integral gives zero contribution in the limit $q \to 1$, while the second one gives a contribution of order $(1-q)^{-1}$. Indeed, the integration over (h, θ) parametrized by q is nothing but an exponential Dirac-delta representation. Whenever the peak of this delta representation lies in the region Ω_ξ, which means that $(h_0, \theta_0) \in \Omega_\xi$ the integral results in a finite contribution. The contributions of order $(1-q)^{-1}$ arises from the points $(h_0, \theta_0) \in \mathcal{C}(\Omega_\xi)$. Therefore, we can write

$$g_1^{RS}(q, \gamma, L) = \left\langle\!\!\left\langle \int_{\mathcal{C}(\Omega_\xi)} \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - l) \ln[1]_\xi(h_0, \theta_0) \right\rangle\!\!\right\rangle_\xi \tag{B1}$$

where we have introduced the shorthand notation

$$[1]_\xi(h_0, \theta_0) = \int_{\Omega_\xi} \frac{dh}{\sqrt{2\pi(1-q)}} \frac{d\theta}{\sqrt{2\pi(1-q)}} \times \exp\left[-\frac{(h-h_0)^2 + (\theta-\theta_0)^2}{2(1-q)} \right]. \tag{B2}$$

Next, for a given $(h_0, \theta_0) \in \mathcal{C}(\Omega_\xi)$ the main contribution arising from the function $[1]_\xi(h_0, \theta_0)$ is obtained for those points $(h, \theta) \in \Omega_\xi$ which minimize the distance $(h - h_0)^2 + (\theta - \theta_0)^2$. To calculate this minimal distance, we split up $\mathcal{C}(\Omega_\xi)$ into three subregions according to fig. 7 in the case of $\xi = 1$

$$\mathcal{R}_1 = \begin{cases} h_0 < \frac{\kappa}{\sqrt{a}} \\ \theta_0 > 0 \end{cases} \tag{B3}$$

$$\mathcal{R}_2 = \begin{cases} \left(h_0 - \frac{\kappa}{\sqrt{a}}\right)\sqrt{\gamma(1-a)} < \theta_0 < 0 \\ h_0 < \frac{\kappa}{\sqrt{a}} \end{cases} \tag{B4}$$

$$\mathcal{R}_3 = \begin{cases} \theta_0 < 0 \\ \frac{\theta_0}{\sqrt{\gamma(1-a)}} + \frac{\kappa}{\sqrt{a}} < h_0 < \frac{\kappa}{\sqrt{a}} - \theta_0\sqrt{\gamma(1-a)}\,. \end{cases} \tag{B5}$$

Computing the minimal distances for such subregions is straightforward and leads to

$$d_{min}^{\mathcal{R}_1} = \left(\frac{\kappa}{\sqrt{a}} - h_0\right)^2 \tag{B6}$$

$$d_{min}^{\mathcal{R}_2} = \left(\frac{\kappa}{\sqrt{a}} - h_0\right)^2 + \theta_0^2 \tag{B7}$$

$$d_{min}^{\mathcal{R}_3} = \frac{1}{1+\gamma(1-a)}\left(\frac{\kappa}{\sqrt{a}} - \theta_0\sqrt{\gamma(1-a)} - h_0\right)^2 \tag{B8}$$

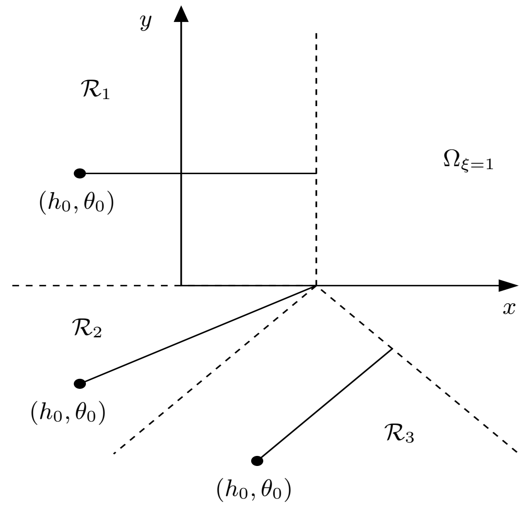

FIG. 7: Schematic representation of the subregions and minimal distances for $\mathcal{C}(\Omega_\xi = 1)$

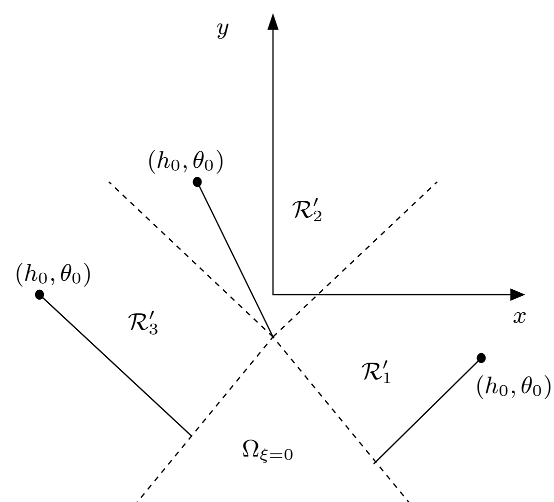

FIG. 8: As figure 7 for $\mathcal{C}(\Omega_\xi = 0)$

By redefining $h_0 - \kappa/\sqrt{a} \to h_0$ and $\gamma' = \sqrt{\gamma(1-a)}$ we recover the expressions (28)-(30).
We proceed analogously for the region $\mathcal{C}(\Omega_{\xi=0})$. We split this region into three subregions as shown in fig. 8

$$\mathcal{R}_1' = \begin{cases} h_0 > 0 \\ -\frac{1}{\sqrt{\gamma(1-a)}}\Big(h_0 + \frac{\kappa}{\sqrt{a}}\Big) < \theta_0 < \sqrt{\gamma(1-a)}h_0 - \frac{\kappa}{\sqrt{\gamma a(1-a)}} \end{cases} \tag{B9}$$

$$\mathcal{R}_2' = \begin{cases} -\frac{1}{\sqrt{\gamma(1-a)}}\Big(\theta_0 + \frac{\kappa}{\sqrt{\gamma a(1-a)}}\Big) < h_0 < \frac{1}{\sqrt{\gamma(1-a)}}\Big(\theta_0 + \frac{\kappa}{\sqrt{\gamma a(1-a)}}\Big) \\ -\frac{\kappa}{\sqrt{\gamma a(1-a)}} < \theta_0 < \infty \end{cases} \tag{B10}$$

$$\mathcal{R}_1' = \begin{cases} h_0 < 0 \\ -\frac{1}{\sqrt{\gamma(1-a)}}\Big(-h_0 + \frac{\kappa}{\sqrt{a}}\Big) < \theta_0 < -\sqrt{\gamma(1-a)}h_0 - \frac{\kappa}{\sqrt{\gamma a(1-a)}}\,. \end{cases} \tag{B11}$$

The minimal distances are given by

$$d_{min}^{\mathcal{R}_1'} = \frac{\big(\sqrt{\gamma a(1-a)}\theta_0 + \kappa + \sqrt{a}h_0\big)^2}{a[1+\gamma(1-a)]} \tag{B12}$$

$$d_{min}^{\mathcal{R}_2'} = h_0^2 + \Big(\frac{\kappa}{\sqrt{\gamma a(1-a)}} + \theta_0\Big)^2 \tag{B13}$$

$$d_{min}^{\mathcal{R}_3'} = \frac{\big(\sqrt{\gamma a(1-a)}\theta_0 + \kappa - \sqrt{a}h_0\big)^2}{a[1+\gamma(1-a)]} \tag{B14}$$

and redefining $\kappa/\sqrt{\gamma a(1-a)}+\theta_0 \to \theta_0$ we find (31)-(33).

APPENDIX C: RS STABILITY

Starting from the stability matrix formed by the second derivatives of Φ (recall eq.(A8)) with respect to the order parameters and the conjugated variables, we find that only transverse fluctuations are relevant.

These transverse fluctuations are characterized by four eigenvalues with degeneracy $n(n-3)/2$, given by the roots of the fourth degree characteristic polynomial $P(\lambda)$

$$P(\lambda) = \begin{vmatrix} \Delta_q - \lambda & \Delta_c & 1 & 0 \\ \Delta_c & \Delta_r - \lambda & 0 & 1 \\ 1 & 0 & \Delta_{\widehat{q}} - \lambda & 0 \\ 0 & 1 & 0 & \Delta_{\widehat{r}} - \lambda \end{vmatrix} = [(\Delta_q - \lambda)(\Delta_{\widehat{q}} - \lambda) - 1][(\Delta_r - \lambda)(\Delta_{\widehat{r}} - \lambda) - 1] - \Delta_c^2(\Delta_{\widehat{q}} - \lambda)(\Delta_{\widehat{r}} - \lambda) \tag{C1}$$

with the coefficients Δ given by

$$\Delta_q = \frac{\alpha}{q^2}\int \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - t)\left\langle\!\!\left\langle \Big\{\frac{\partial^2}{\partial h_0^2}\ln[1]_\xi(h_0,\theta_0)\Big\}^2 \right\rangle\!\!\right\rangle_{\xi_o} \tag{C2}$$

$$\Delta_r = \frac{\alpha}{r^2}\int \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - t)\left\langle\!\!\left\langle \Big\{\frac{1}{\gamma}\frac{\partial^2}{\partial \theta_0^2}\ln[1]_\xi(h_0,\theta_0)\Big\}^2 \right\rangle\!\!\right\rangle_{\xi_o} \tag{C3}$$

$$\Delta_c = \frac{\alpha}{qr}\int \mathcal{D}(h_0)\mathcal{D}(\sqrt{\gamma}\theta_0 - t)\left\langle\!\!\left\langle \Big\{\frac{1}{\sqrt{\gamma}}\frac{\partial^2}{\partial h_0 \partial \theta_0}\ln[1]_\xi(h_0,\theta_0)\Big\}^2 \right\rangle\!\!\right\rangle_{\xi_o} \tag{C4}$$

$$\Delta_{\widehat{q}} = (1-q)^2 \tag{C5}$$

$$\Delta_{\widehat{r}} = (1-r)^2 = \gamma^2(1-q)^2 \tag{C6}$$

where we recall that $(1-r) = \gamma(1-q)$ and the function $[1]_\xi(h_0,\theta_0)$ is defined in (B2).

Next, the limit $q \to 1$ has to be taken. Using the asymptotic expansion of $[1]_\xi(h_0,\theta_0)$ discussed in appendix B we can compute the asymptotic behavior of the coefficients Δ_q, Δ_r and Δ_c. After a lot of algebra we

finally arrive at the expressions (39)-(43) with the integration regions and minimal distances given by (22)-(33). In this limit, it turns out that an analytical expression can be found for the eigenvalues. First, we notice that the determinant of the matrix remains finite in the limit. Since the determinant is the product of the eigenvalues, it follows that this product is finite. Two possibilities arise, either all eigenvalues are finite, or two of them tend to zero and two to infinity with the same ratio. It is not hard to prove that the first choice is incorrect. Hence, two of the eigenvalues have to behave asympotically as $(1-q)^{\pm n}$. One can check that only $n=2$ is possible. This allows us to split $P(\lambda)$ into two polynomials which give the solutions around zero and around infinity. These polynomials read

$$\begin{aligned} P_0(\lambda) &= [\Delta_q(\Delta_{\widehat{q}}-\lambda)-1][\Delta_r(\Delta_{\widehat{r}}-\lambda)-1] \\ &\quad -\Delta_c^2(\Delta_{\widehat{q}}-\lambda)(\Delta_{\widehat{r}}-\lambda) \\ P_\infty(\lambda) &= (\Delta_q-\lambda)(\Delta_r-\lambda)-\Delta_c^2\,. \end{aligned} \tag{C7}$$

From these two polynomials the four eigenvalues (35)-(38) can be found. We remark that in the limit $a \to 1$ we find back the stability criteria for the original Gardner capacity problem.

[1] D. R. Dominguez Carreta and E. Korutcheva, Phys. Rev. E **62**, 2620 (2000).
[2] D. Bollé and T. Verbeiren, Phys. Let. A **297**, 156 (2002).
[3] J. Hertz, A. Krogh and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City 1991).
[4] B. Müller, J. Reinhardt and M. T. Strickland, *Neural Networks: An Introduction* (Springer, Berlin 1995).
[5] M. Blume, V.J. Emery and R.B. Griffiths, Phys. Rev. A, **4**, 1071 (1971); M. Blume, Phys. Rev., **141**, 517 (1966); H.W. Capel, Physica **32**, 966 (1966).
[6] J. M. de Araújo, F. A. da Costa and F. D. Nobre, Eur. Phys. J. B **14**, 661 (2000).
[7] M. Opper and W. Kinzel, *Models of Neural Networks III* ed. E. Domany, J. L. van Hemmen and K. Schulten, 151 (Springer, New-York 1996).
[8] D. Saad, ed.*On-line Learning in Neural Networks* (Cambridge University Press 1998).
[9] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press 2001)
[10] D. Bollé, J. Busquets Blanco and G. M. Shim, cond-mat/0206569, to appear in Physica A.
[11] J. P. Nadal and A. Rau, J. Phys. I France **1**, 1109 (1991).
[12] F. Gerl and U. Krey, J. Phys. A **27**, 7353 (1994).
[13] D. A. Köhring, J. Phys. France **51**, 145 (1990).
[14] S. Mertens, H. M. Köhler and S. Bös, J. Phys. A **24**, 4941 (1991).
[15] D. Bollé, P. Dupont and J. van Mourik, Europhys. Lett. **15**, 893 (1991).
[16] D. Bollé, R. Kühn and J. van Mourik, J. Phys. A **26**, 3149 (1993).
[17] D. Bollé and R.Erichsen, Jr, Phys. Rev. E **59**, 3386 (1999)
[18] D. Bollé and P.Kozlowski, Phys. Rev. E **64**, 011915 (2001).
[19] E. Gardner, Europhys. Lett. **4**, 481 (1987); J. Phys. A, **21**, 257 (1988).
[20] E. Gardner and B. Derrida, J. Phys. A, **21**, 271 (1988).
[21] F. Gerl, K. Bauer and U. Krey, Z. Phys. B **88**, 339 (1992).
[22] J.R. de Almeida and D. Thouless, J. Phys. A, **11**, 983 (1978)
[23] M. Mézard, G. Parisi and M.A. Virasoro, *Spin Glass Theory and Beyond* (Singapore, World Scientific, 1987)
[24] D. Bollé and J. van Mourik, J. Phys. A **27**, 1151 (1994).